What to do with your raw uBiome data

For most people who do just one or two tests, the standard uBiome web page (at <http://app.ubiome.com)> offers plenty of information. You can look at the percentage breakdown of different bacteria, compare them with other users or to yourself over time, and dig deeper with descriptions of the most common organisms and what they do. But if you really want to understand your microbiome, uBiome offers much more: full access to all the raw data: literally millions of snippets of genetic information ready to analyze.

I recently wrote a detailed description for the July 2015 issue of O’Reilly’s Biocoder magazine (available as a free download here: <http://www.oreilly.com/biocoder/> ) and I encourage you to read the whole thing for more details, but here’s a short summary of three steps to get more from your data:

# STEP 1

First, click the “Download taxonomy” button on [the web page for your sample](https://app.ubiome.com/).

Although it will look like gobbley-gook, you can turn this into an Excel spreadsheet easily enough: just select the info on the page and copy/paste it into a site that will convert it automatically into a CSV file. (I use <http://konklone.io/json/> or <http://www.convertcsv.com/json-to-csv.htm)> Read the CSV file into Excel and there will be three columns you care about: “tax\_rank”, “tax\_name” and “count\_norm”.

Then it’s a simple matter of running some standard Excel filtering operations on the data. Filter tax\_rank by “Phylum” and then sort the count\_norm field from largest to smallest. The count\_norm numbers correspond to parts per million – just divide by 10,000 to get the percentages.

By the way, a big bonus awaits you in the taxonomy file that you can’t get from the standard web page: species information. [Most scientists trust](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2045242/) the 16S rRNA technology down to the genus level, but there is more uncertainty at the species level, so uBiome doesn’t publish it to the web page. Drag it into Excel, though, and you can make up your own mind about whether you trust the species info or not.

# STEP 2

Zip to the new uBiome open source microbiome-tools GitHub page and download ubiomeCompare.py. If you have the Python language on your computer (all Macs come with it built-in), you can run this file without installing anything extra. If not, download it for free at <http://python.org>

Let’s say your spouse has the sample in a file called Wife1.JSON and yours is in Husband1.JSON. On a Mac, open Terminal and run the following command:

> python ubiomeCompare.py –u Husband1.JSON Wife1.JSON > HusbandUnique.CSV

The new file, HusbandUnique.CSV contains just those organisms that are unique to the Husband1 sample, i.e. are found in the husband’s biome and not the wife’s.

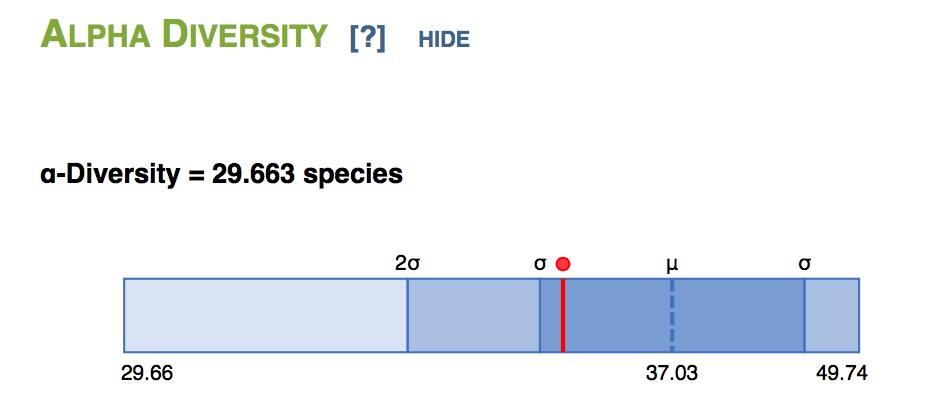
Similarly, the following command will give you a file that contains the relative differences between every organism in Husband1 and Wife1:

> python ubiomeCompare.py –c Husband1.JSON Wife1.JSON > HusbandUnique.CSV

# STEP 3

Finally, if you’re really into serious number crunching, uBiome gives you the raw output from their expensive Illumina NextSeq 500 in the form of FASTQ files. If you know what that is, you probably already know how to read them, but if not please look at the BioCoder article for an introduction. With a little work, the FASTQ files will let you see precisely which genes were detected in your sample. Since so much of the microbiome is still unexplored, you may find pieces that are missing from the regular uBiome output, so this is your chance to go straight to the underlying genetic information for more.

For example, I was able to compute the following measure of diversity from my most recent sample. It’s a measure I’ll track for all of my samples



# Going Even Further

I’ve barely scratched the surface of what’s possible when you use your raw uBiome data. Please look at the BioCoder article for more step-by-step instructions, and [contact me](http://twitter.com/sprague) if you have other questions. Meanwhile, please upload your own samples to the uBiome GitHub and feel free to compare yours to the ones there, so we can all learn more together.